

---

## ***2014 Emissions Modeling Platform Spatial Surrogate Documentation***

**Contract No:** EP-D-12-044  
**Work Assignment:** 5-02 (Platform Support)  
**Version No:** 1.0

---

**Prepared for:** Alison Eyth  
U.S. EPA OAQPS  
C339-02 USEPA Mailroom  
Research Triangle Park, NC 27711

**Prepared by:** Zachariah Adelman  
Institute for the Environment  
The University of North Carolina at Chapel Hill  
100 Europa Dr. Suite 490  
Chapel Hill, NC 27517-1105

**Date due:** October 1, 2016

# Contents

<b>Acknowledgements .....</b>	<b>4</b>
<b>1 Introduction.....</b>	<b>4</b>
<b>2 Shapefile Database.....</b>	<b>7</b>
2.1 U.S. Census-based Surrogates .....	7
2.1.1 Population and Housing .....	7
2.1.2 Home Heating Fuels .....	8
2.2 Transportation Surrogates .....	8
2.2.1 Roadways .....	8
2.2.2 Rail Length .....	10
2.2.3 Rail Freight Density .....	10
2.2.4 Rail Yards.....	11
2.2.5 Bus Terminals.....	12
2.2.6 Extended Idle Locations .....	12
2.3 Shipping and Port Surrogates.....	15
2.3.1 Waterways .....	15
2.3.2 Marine Ports .....	16
2.3.3 Gulf of Mexico Platforms.....	16
2.3.4 Offshore Commercial Shipping.....	17
2.4 FEMA Building Footprints .....	18
2.5 Landcover Surrogates .....	18
2.6 Oil and Gas Production.....	20
2.6.1 Eastern Research Group, Inc. Oil and Gas Production Data (HPDI Database) .....	20
2.7 Other Industrial and Commercial Activities .....	22
2.7.1 Refineries and Tank Farms.....	22
2.7.2 Mines .....	23
2.7.3 Commercial Timber .....	23
2.7.4 Golf Courses .....	24
2.7.5 Airport Point Emissions .....	24
2.7.6 Airport Area .....	25

<b>3</b>	<b>Known Issues and Future Work .....</b>	<b>25</b>
<b>4</b>	<b>Surrogate Revision History .....</b>	<b>26</b>
<b>5</b>	<b>REFERENCES .....</b>	<b>26</b>
	<b>Appendix A – Surrogate Specifications .....</b>	<b>28</b>

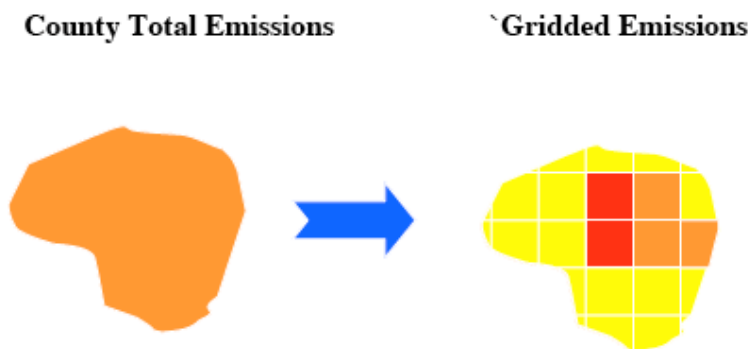
## Acknowledgements

Eastern Research Group, Inc. (ERG) contributed to this work through the development of the Gulf of Mexico platform and support vessel density Shapefile and through the development of oil and gas production Shapefiles and Surrogates.

## 1 Introduction

This memo describes the collection, processing, and development of updated geospatial data for calculating spatial surrogates to be used in EPA emissions modeling platforms.

Spatial surrogates are used to allocate county level emission inventories into the rectangular grid cells used by urban and regional scale air quality models. Spatial surrogates are based on data at resolutions different from the county-level, such as census tracts or road locations, and can therefore be used to allocate the emissions more specifically than assuming they are uniformly distributed throughout the county. For example, motor vehicle emissions from interstate highways can be placed into the grid cells that intersect the highways themselves, and dry cleaning emissions can be allocated to areas of the county that have higher population, or the areas that have more dry cleaners - if data are available on specific locations of dry cleaners. Spatial surrogates consist of values between 0 and 1.0 that specify the fraction of the county emissions that should be allocated to each grid cell that intersects the county. The spatial surrogate fractions for each county sum to 1.0, except for counties that intersect the edge of a modeling grid and are not completely contained within the modeling domain. The emissions for a specific county within a grid cell are computed by multiplying the surrogate fraction for that county and grid cell by the total emissions for the county (Eyth et al., 2007). Figure 1 illustrates how spatial surrogates impact the emissions levels within the grid cells of a county.



**Figure 1. Spatial surrogate schematic**

Spatial surrogates have three components:

1. Weight Shapefile – geospatial dataset used for allocating (or weighting) county-level emission inventory sources to a modeling grid. Examples include roadways, population density, or the point locations of marine ports. This memo describes the weight Shapefile updates completed under this Work Assignment.
2. Data Shapefile – geospatial dataset that corresponds to the administrative unit boundaries of the emission inventory. For this Work Assignment, the data Shapefile defines U.S. county boundaries. EPA specified that the year 2014 county boundaries from the U.S. Census be used as the data Shapefile for this Work Assignment.
3. Output grid or polygon – definition of the modeling grid(s) or polygon(s) upon which the spatial surrogates map emission inventories. For this Work Assignment, two output grids include continental U.S. modeling grids at resolutions of 12-km and 4-km.

The U.S. EPA surrogate database was last updated in 2014 for use with the 2011 National Emissions Inventory (UNC-IE, 2014). Those surrogates incorporated new data for most of the weight shapefiles and included several completely new surrogates. Under the current work assignment, UNC-IE updated the spatial surrogates for use with the 2014 National Emissions Inventory (NEI). These updates include the use of 2014 county boundaries and new weight data that have become available since the 2011 NEI surrogate updates. The data and configurations used to update these surrogates are described here. The surrogates described in this memo were generated using a combination of the Surrogate Tool<sup>1</sup> of the Spatial Allocator (SA)<sup>2</sup> and a new, forthcoming PostgreSQL/PostGIS utility.

UNC-IE generated spatial surrogates on the CONUS domain at resolutions of 12-km and 4-km. All of the surrogates were generated on a Lambert Conformal Conic projection (lat<sub>0</sub> = 40° north, lon<sub>0</sub> = 97° west, true\_lat1 = 33° north, true\_lat2 = 45° north) assuming a spherical earth with a radius of 6,370,000 m. The grid definitions of the three surrogate domains are provided in Table 1.

**Table 1. Surrogate domain definitions**

Parameter	12km	4km
X-orig	-2,736,000 m	-2,736,000 m
Y-orig	-2,088,000 m	-2,088,000 m
columns	444	1332
rows	336	1008

UNC-IE used the gapfilling, normalization, and quality assurance (QA) tools of the SA to ensure completeness and quality of the surrogates. Normalizing the surrogates eliminates surrogate values that do not equal 1.0, which can result from rounding errors in the surrogate calculations. The QA steps check for surrogates that do not equal 1.0, missing surrogates, and surrogates that are gapfilled. In addition to these tabulated QA summaries, UNC-IE produced spatial plots of the surrogates on each of the CONUS domains. The QA tables, spatial plots, and a README file

<sup>1</sup> [http://www.cmascenter.org/sa-tools/documentation/4.0/html/srgtool/SurrogateToolUserGuide\\_4\\_0.htm](http://www.cmascenter.org/sa-tools/documentation/4.0/html/srgtool/SurrogateToolUserGuide_4_0.htm)

<sup>2</sup> <http://www.cmascenter.org/sa-tools/>

describing the surrogates are packaged in a “QA” directory in the 2014 NEI spatial surrogate Linux tar archives. The surrogates, shapefiles, and accompanying QA data are available from the U.S. EPA CHIEF<sup>3</sup> website as part of the OAQPS 2014 emissions modeling platform and from the CMAS Center Data Clearinghouse<sup>4</sup>.

UNC-IE collected, created, and prepared several GIS shapefiles for generating the 2014 spatial surrogates. The American Community Survey (ACS) of the U.S. Census includes annual 5-year average estimates that are released between the decadal census. The 2010-2014 ACS (U.S. Census Bureau, 2014) provided the basis for the population and housing surrogates, including the home heating surrogates. Year 2013 roadway data modeling onroad mobile sources came from the U.S. Department of Transportation Highway Performance Monitoring System (HPMS; U.S. DOT, 2013). Rail data came from the 2014 National Transportation Atlas Database (NTAD; U.S. BTS, 2014) and the U.S. Census 2014 TIGER/Line Files (U.S. Census Bureau, 2014). Landcover data from the 2011 National Landcover Database (NLCD; Homer et al., 2015) were used to create several surrogates related to built and natural features. The FEMA HAZUS-MH v2.0, which was released in September 2011, contains building square footage data that were used to identify the locations of different types of built features. The U.S. Forest Service 2016 Planned and Accomplished Timber Harvest was used to create a surrogate for logging equipment and the 2011 USGS Mineral Resources System was used to update the mining equipment surrogate (USGS, 2005). Year 2015 data on energy system commodities from the Energy Information Administration (EIA) were used to create a new surrogate for refineries and tank farms. The Points of Interest (POI) Factory, public golf course shapefile was used to create a new golf course surrogate for use with commercial turf maintenance equipment. Eastern Research Group, Inc. aggregated basin-level HPDI production and activity data to the continental scale for upstream oil and gas sources (Reference).

Several new surrogates were created to support the processing of MOVES mobile sources. The Urban/Rural Restricted/Unrestricted road miles and Annual Average Daily Traffic (AADT) data from HPMS were processed using Federal Highway Administration (FHWA) urban and rural distinctions<sup>5</sup> that are consistent with the roadway classifications in MOVES. UNC-IE updated the Extended Idle Locations shapefile that is used for MOVES rate per hour (RPH) sources.

The NLCD data were leveraged to create several new surrogates that are based on different levels of development intensity. UNC-IE worked with the U.S. EPA on identifying surrogates that reached a middle ground between the non-specific total population (100) surrogate and the overly prescriptive FEMA building square footage surrogates (5\*\*). In creating combinations of the NLCD development intensity landcover categories (e.g. low+medium intensity), we tried to capture the locations of non-point emissions activities in a way that does not concentrate emissions into false hot spots.

---

<sup>3</sup> <http://www.epa.gov/chief/>

<sup>4</sup> <http://www.cmascenter.org/download/data.cfm>

<sup>5</sup>

[https://www.fhwa.dot.gov/planning/processes/statewide/related/highway\\_functional\\_classifications/section06.cfm](https://www.fhwa.dot.gov/planning/processes/statewide/related/highway_functional_classifications/section06.cfm)

The list of GIS Shapefiles used for the updated spatial surrogates and details of how these data were preprocessed for generating surrogates are provided in Section 2. A Google Spreadsheet (<https://docs.google.com/spreadsheets/d/1byvKzcSmuHmzjbRYE8KTmhhAIH2oY9OjKtYBx5eILFg/edit?usp=sharing>) that accompanies this document includes the detailed specifications of the Shapefiles and surrogates that were collected and updated under this Work Assignment. All of the shapefiles used to develop the 2014 spatial surrogates were processed in their native projections.

## 2 Shapefile Database

### 2.1 U.S. Census-based Surrogates

#### 2.1.1 Population and Housing

The 2010-2014 TIGER/Line database contains American Community Survey (ACS) 5-year estimates of population and housing unit counts at the Census block level for each state. UNC-IE calculated urban and rural areas using Census block groups. Urban was defined as Census block groups that have a population density of at least 1,000 people per square mile and everything else was defined as rural. In addition, Census block groups are classified based on the number of housing units per square kilometers (housing unit density [HUD]) and the attribute HUD\_CODE is assigned to census block groups with the following conditions:

1. Urban: HUD  $\geq$  1000
2. Suburban: HUD  $\geq$  125 and HUD  $<$  1000
3. Exurban: HUD  $\geq$  6 and HUD  $<$  125
4. Rural: HUD  $<$  6

The attributes in the file used to compute surrogates include:

- POP2014 – 2014 population
- HU2014 – 2014 housing units
- HUD\_CODE – census block group codes; 1=urban, 2=suburban, 3=exurban, 4=rural

*Shapefile Name:* ACS\_2014\_5HR\_PopHousing

*Shapefile Type:* Polygon

*Year:* 2010-2014 5-year average

*Spatial Coverage:* All U.S. States and territories

*Attributes:* STATEFP,COUNTYFP,TRACTCE,BLKGRPCE,GEOID,NAMELSAD,MTFCC,FUNCSTAT,ALAND,AWATER,INTPTLAT,INTPTLON,Shape\_Leng,Shape\_Area,GEOID\_Data,FIPSSTCO,HU2014,UTIL\_GAS,LP\_GAS,ELEC,FUEL\_OIL,COAL,WOOD,SOLAR,TOTAREA,POPDENS,HUDENS,POP2014,HUCH14\_10,POPCH14\_10,POP2010,HU2010,POP2000,HU2000,POPD10,HUD14\_KM2,HUD\_CODE,UACE10,GEOID10,NAME10,LSAD10,MTFCC10,UATYP10,ALAND10,AWATER10,POPD\_CODE

*Projection:* Geographic

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 100, 110, 131, 132, 134, 137, 140

### **2.1.2 Home Heating Fuels**

The American Community Survey (ACS) is a U.S. Census project that collects yearly demographic and housing information from randomly selected households throughout the U.S. The data are aggregated in 5 and 10-year increments to provide Census tract estimates for the statistics collected during the survey. Home heating fuel type from the ACS was used to develop spatial surrogates for home heating sources (i.e. residential wood combustion) in the nonpoint inventory. The ACS 5-year 2014 survey results that were released in 2015 represent data collected from 2001-2014. These data were used to create home heating surrogates.

The attributes in the file used to compute surrogates include:

- UTIL\_GAS - number of housing units using Utility Gas for primary heating
- WOOD - number of housing units using Wood for primary heating
- FUEL\_OIL - number of housing units using Fuel Oil for primary heating
- COAL - number of housing units using Coal for primary heating
- LP\_GAS - number of housing units using Bottled Gas for primary heating

*Shapefile Name:* ACS\_2014\_5HR\_PopHousing

*Shapefile Type:* Polygon

*Year:* 2010-2014 5-year average

*Spatial Coverage:* All U.S. States and territories

*Attributes:* STATEFP,COUNTYFP,TRACTCE,BLKGRPCE,GEOID,NAMELSAD,MTFCC,FUNCSTAT,ALAND,AWATER,INTPTLAT,INTPTLON,Shape\_Leng,Shape\_Area,GEOID\_Data,FIPSSTCO,HU2014,UTIL\_GAS,LP\_GAS,ELEC,FUEL\_OIL,COAL,WOOD,SOLAR,TOTAREA,POPDENS,HUDENS,POP2014,HUCH14\_10,POPCH14\_10,POP2010,HU2010,POP2000,HU2000,POPD10,HUD14\_KM2,HUD\_CODE,UACE10,GEOID10,NAME10,LSAD10,MTFCC10,UATYP10,ALAND10,AWATER10,POPD\_CODE

*Projection:* Geographic

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 150, 160, 170, 180, 190

## **2.2 Transportation Surrogates**

### **2.2.1 Roadways**

UNC-IE aggregated state-level 2013 HPMS Shapefile data to create a national file with urban and rural roadways. UNC-IE projected the data to a U.S. national Lambert Conformal Conic projection on a normal sphere. FHWA urban and rural roadway classifications encoded in the HPMS data were used to distinguish between urban and rural roadways. The HPMS F\_System codes (Table 2) were used to distinguish the different road types (DOT, 2013).



**Table 2. HMPS F\_System codes**

Code	Description	
	Rural	Urban
1	Interstate	Interstate
2	Other Principal Arterial	Other Freeways and Expressways
3	Minor Arterial	Other Principal Arterial
4	Major Collector	Minor Arterial
5	Minor Collector	Collector
6	Local	Local

The MOVES 2014 road types (EPA, 2016: Table 2-4) (Table 3) were then mapped to the HPMS F\_System codes using the FHWA urban/rural codes that are also included in the HPMS database.

**Table 3. MOVES2014 road type codes.**

roadTypeID	Description	FHWA Functional Types
1	Off Network	Off Network
2	Rural Restricted Access	Rural Interstate
3	Rural Unrestricted Access	Rural Principal Arterial, Minor Arterial, Major Collector, Minor Collector & Local
4	Urban Restricted Access	Urban Interstate & Urban Freeway/Expressway
5	Urban Unrestricted Access	Urban Principal Arterial, Minor Arterial, Collector & Local
6	Rural Restricted without Ramps	
7	Urban Restricted without Ramps	
8	Rural Restricted only Ramps	
9	Urban Restricted only Ramps	
100	Nonroad	

We created the road type (RDTYPE) attribute in the HPMS shapefiles to use for creating surrogates for the MOVES onroad mobile emissions data. The attributes in the HPMS shapefile used to compute surrogates include:

RDTYPE	Description	HPMS F_System Code
2	Rural Restricted	1 + rural
3	Rural Unrestricted	2, 3, 4, 5 + rural
4	Urban Restricted	1, 2 + urban
5	Urban Unrestricted	3, 4, 5 + urban

An additional attribute for annual average daily traffic (aadt) is used to weight the road types by average activity on each road link.

*Shapefile Names:* hpms2013\_2014\_roads\_surrogate\_rdtype2,  
hpms2013\_2014\_roads\_surrogate\_rdtype3, hpms2013\_2014\_roads\_surrogate\_rdtype4,  
hpms2013\_2014\_roads\_surrogate\_rdtype4

*Shapefile Type:* Line

*Year:* 2013-2014

*Spatial Coverage:* Conterminous U.S.

*Attributes:*

*Projection:* Geographic

*Datum:* NAD83

*Spheroid:*

*Derived Surrogates:* 201, 202, 211, 212, 221, 222, 231, 232, 239, 240, 241, 242, 243, 244

## 2.2.2 Rail Length

UNC-IE used the national 2014 TIGER/Line Shapefile data to create a national railway network file. The TIGER/Line MTFCC codes R1052, R1051, and R1011 were used to define the different classes of rail lines.

The attributes in the file used to compute surrogates include:

- MTFCC – rail classes; 1011=class 1; 1051=class 2; 1052=class 3

*Shapefile Name:* TIGER\_2014\_Rail

*Shapefile Type:* Line

*Year:* 2014

*Spatial Coverage:* All U.S. States and territories

*Attributes:* LINEARID, FULLNAME, MTFCC

*Projection:* Geographic

*Datum:* NAD83

*Spheroid:* GRS80

*Derived Surrogates:* 260, 280

## 2.2.3 Rail Freight Density

UNC-IE used the 2014 National Transportation Atlas Database (NTAD) freight densities to weight the line Shapefile of rail activity locations. The attribute DENS defines freight densities in Millions of Gross Tons (MGT) for each rail link. The original NTAD database includes freight density categories with the attribute DEN11CODE. UNC-IE mapped the density categories to integer densities for use in the surrogate calculations. Table 4 shows the mapping from the DEN11CODE categories to DENS.

**Table 4. NTAD rail density code mapping**

DEN11CODE Category	DENS (MGT)
Unknown	1
<5 MGT	5
5-10 MGT	10
10-50 MGT	25
50-100 MGT	75

>100 MGT	100
----------	-----

The NTAD RAILTYPE codes were used to define the different classes of rail lines.

The attributes in the file used to compute surrogates include:

- DENS – rail link freight density (MGT); see Table 4
- RAILTYPE – rail classes; 1=class 1,2,3 rail; 2=Amtrak; 3=Commuter rail

*Shapefile Name:* NTAD\_2014\_Rail

*Shapefile Type:* Line

*Year:* 2014

*Spatial Coverage:* All U.S. States, no territories

*Attributes* FRAARCID,MILES,STATEAB,STATEFIPS,CNTYFIPS,STCNTYFIPS,FRAREGIONS,RROWNER1,RROWNER2,RROWNER3,TRKRGHTS1,TRKRGHTS2,TRKRGHTS3,TRKRGHTS4,TRKRGHTS5,TRKRGHTS6,TRKRGHTS7,TRKRGHTS8,TRKRGHTS9,STRACNET,SIGSYS,TRACKS,SUBDIV,FRFRANODE,TOFRANODE,NET,PASSNGR,YARDS,INT\_TYPE,DEN11CODE,DENS,RAILTYPE

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* 261, 271, 272, 273

## 2.2.4 Rail Yards

Michelle Bergin and Byeong Kim of GA DNR provided a spreadsheet of rail yard locations and emissions collected as part of the Eastern Regional Technical Advisory Committee (ERTAC) railway inventory development and improvement effort (Bergin et al., 2009). UNC-IE converted this spreadsheet to a point Shapefile for generating a rail yard surrogate for the U.S. The surrogate is based on the sum of the NO<sub>x</sub> and PM<sub>2.5</sub> emissions at the rail yard locations in the ERTAC rail yard inventory.

The attributes in the file used to compute surrogates include:

- NOXPMTOT – total of NO<sub>x</sub> and PM<sub>2.5</sub> emissions at each rail yard

*Shapefile Name:* ERTAC\_railyard\_WRF

*Shapefile Type:* Point

*Year:* 2012

*Spatial Coverage:* All U.S. States, no territories

*Attributes* OBJECTID, OWNER1, OWNER2, OWNER3, STATE, FIPS, YARDNAME, LON, LAT, NOX, PM25, NOXPMTOT,

*Projection:* Lambert Conformal Conic (X<sub>0</sub>=-97.0, Y<sub>1</sub>=33.0, Y<sub>2</sub>=45.0, Y<sub>0</sub>=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 275

## 2.2.5 Bus Terminals

2014 U.S. locations for transportation terminals from the National Transportation Atlas Intermodal Passenger Connectivity Database. This Shapefile was used to define the locations of intercity and transit bus stations for allocating off-network MOVES emissions to modeling grids. This point Shapefile of transportation terminals was obtained from the online National Transportation Atlas Database.

The attributes in the file used to compute surrogates include:

- BUS\_INTERC – point locations of intercity bus terminals
- BUS\_TRANSI – point locations of transit bus terminals

*Shapefile Name:* NTAD\_2014\_ipcd

*Shapefile Type:* Point

*Year:* 2014

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* Point\_ID, CITY, STATE, METRO\_AREA, Lon, Lat, FERRY\_TRAN, FERRY\_INTE, BUS\_TRANSI, BUS\_INTERC, BUS\_CODE, BUS\_SUPPLE, RAIL\_INTER, RAIL\_COMMU, RAIL\_HEAVY, RAIL\_LIGHT, AIR\_SERVICE, INTERCITY\_, TRASIT\_SE, CBSA\_CODE, CBSA\_TYPE, MODES\_SRV, MODE\_BUS, MODE\_AIR, MODE\_RAIL, MODE\_FERRY

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* 258, 259

## 2.2.6 Extended Idle Locations

Locations of extended idle emissions for tractor-trailers. The base parking area shapefile was provided by EPA and it included 7 different classes of parking locations:

1. State Department of Transportation (DOT) visitor centers
2. DOT welcome centers
3. DOT rest areas
4. DOT weigh stations
5. DOT parking areas
6. Private truck stops
7. Private retail locations, including Walmart and McDonalds

Along with the latitude-longitude coordinates of the parking locations, the database includes attributes indicating the number of parking spots at each location to use for weighting a spatial

surrogate. The attributes for the number of parking spots for the different types of locations were pulled from multiple data sources. The EPA database included the number of parking spots for many of the private truck stops but did not include the number of spots for most of the DOT or private retail locations. The following steps were followed to mine parking spot data to develop a complete database of extended idle locations.

#### State DOT Locations

Nationwide request from all state DOTs for the number of parking spots at state-run locations returned data from the WV, IN, MO, WI, NC, CT, ME, UT, VA, GA, and FL. Many of these states reviewed the default data pulled from the web and updated these data with new information. We reconciled the data provided by the states with the national data provided by EPA. Manual matching was done using Bing satellite imagery for locations in which the parking area exit ramp coordinates were provided rather than the actual parking lot location. The matching yielded a fairly complete database of coordinates and the number (or range) of truck parking spots at DOT locations.

#### Private Truck Stops

Data downloaded from Pilot, Petro, and TA truck stop company websites include the coordinates and number of truck parking spots at all locations nationally. Merged these data with the EPA database using a 300 meter buffer to exclude duplicates. These data appeared more reliable than the number of parking spot attributes in the EPA database based on manual comparisons with Bing satellite imagery.

#### Private Retail Locations

The EPA database included coordinates and number of parking spots for some retail and food service areas. Walgreens composed the largest number of private retail locations in the database and did not include any information on the number of truck parking spots at any of the locations.

#### Gapfilling the Number of Parking Spots

After creating an inventory of the truck parking spot locations from the collected data and merging these data into the EPA database, we used the following rules to fill in the number of truck parking spots:

- If number of spots listed as “<20”, set to 20
- If number of spots listed as “20-69”, set to 45
- If number of spots listed as “>70”, set to 70
- If retail locations listed the number of spots as “unknown”, set to 2
- Weigh stations and parking areas all listed the number of spots as “unknown”, set to 2
- For rest areas with missing truck parking spots, calculated 1<sup>st</sup> quartile, median, and 3<sup>rd</sup> quartile from all known DOT rest area data and used these as low, medium, and high attributes for the number of spots; results: low = 14, median = 18, high = 31
- For truck stops with missing truck parking spots, calculated 1<sup>st</sup> quartile, median, and 3<sup>rd</sup> quartile from all known truck stop data and used these as low, medium, and high attributes for the number of spots; results: low = 80, median = 123, high = 188

#### Develop Shapefile

Created a shapefile from the modified/gapfilled EPA database by converting the point locations to an NAD83 datum and Lambert Conformal Conic projection. The shapefile includes the attributes truckpark\_low, truckpark\_med, and truckpark\_high to use for weighting spatial surrogates. If the number of truck parking spots at a location was known in the database and there was no need to gapfill, these low/med/high attributes were set to all equal the number of spots at the location.

#### Updates for 2016

Several updates to the extended idling location shapefile were incorporated for the 2014 modeling surrogates:

- Idling location data from the Maricopa Association of Governments were used to replace all of the previous data for Maricopa County, AZ
- The coordinates for one truck stop in St. Louis County, MO were corrected to place truck stop in the correct location
- Locations for the Loves Truck Stop chain were added to the database
- Locations for public and private truck stops in the Chicago area, including state travel plazas, were added to the database.
- Missing Walmart locations were added to the database.

The attributes in the shapefile used to compute surrogates include:

- trucks\_med – median number of truck parking spaces at potential idling locations

*Shapefile Name:* 2016\_PIL\_16Aug

*Shapfile Type:* Point

*Year:* 2016

*Spatial Coverage:* Conterminous U.S.

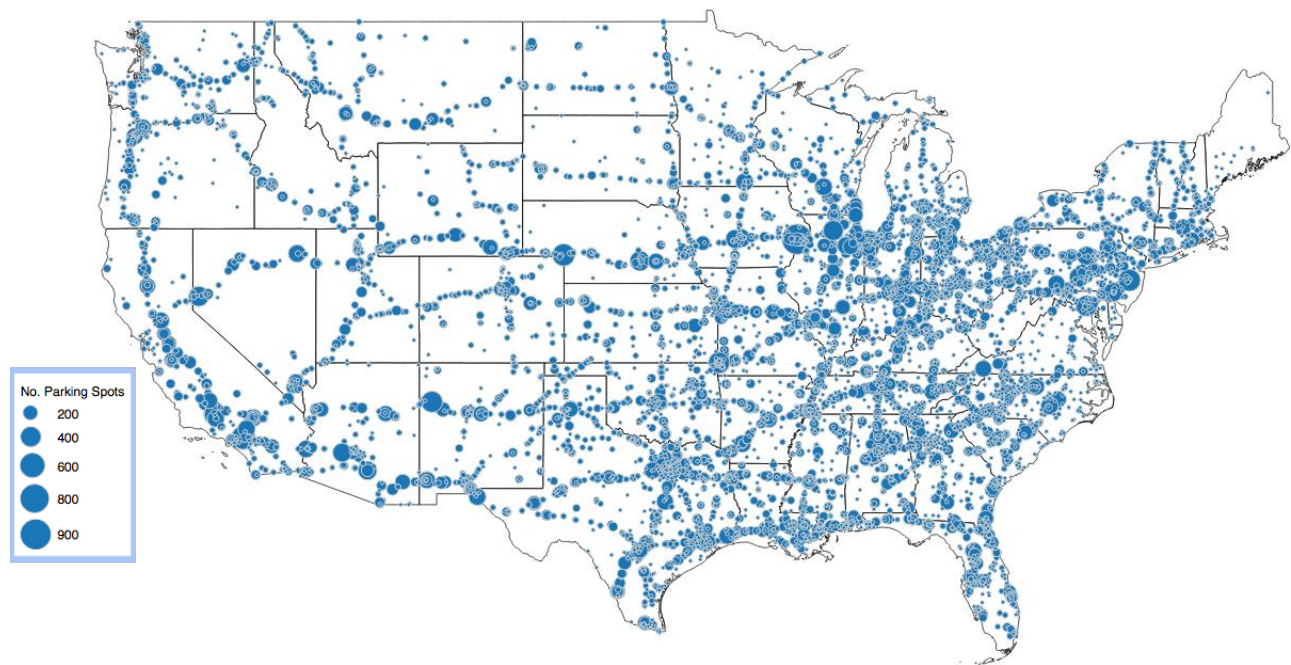
*Attributes:* truck\_park, name, address, city, state, zip, latitude, longitude, site\_cat, source, directions, alt\_source, trucks\_low, trucks\_med, trucks\_high

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates: 205*



**Figure 2. 2016 Potential idling locations**

## **2.3 Shipping and Port Surrogates**

### **2.3.1 Waterways**

2014 National Transportation Atlas Database (NTAD) data for waterway lengths.

The attributes in the file used to compute surrogates include:

- LENGTH – navigable waterway lengths

*Shapefile Name:* NTAD\_2014\_Waterway

*Shapefile Type:* Line

*Year:* 2011

*Spatial Coverage:* All U.S. states and territories

*Attributes:* LENGTH, FEATUREID, ANODE, BNODE, LINKNAME, RIVERNAME, AMILE, BMILE, LENGTH1, LENTH\_SRC, LINKTYPE, CTRL\_DEPTH, WATERWAY, GEO\_CLASS, FUNC\_CLASS, WTWY\_TYPE, CHART\_ID, NUM\_PAIRS, WHO\_MOD, DATE\_MOD, HEADING, STATE, FIPS, FIPS2, NONUS, VERSION

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* 807

### **2.3.2 Marine Ports**

2014 marine port areas weighted by NEI2014 activity in KW.

The attributes in the file used to compute surrogates include:

- Area\_sqmi – port areas in miles<sup>2</sup>
- ACTIVITYKW – annual KW activities for port vessels from the NEI2014

*Shapefile Name:* Ports\_2014NEI

*Shapefile Type:* polygon

*Year:* 2014

*Spatial Coverage:* All U.S. states and territories

*Attributes:* XLongitude,YLatitude,ShapeID,FIPS,NAME,Area\_sqmi,PortID,ShapeType,ACTIVITYKW,TYPE

*Projection:* Equidistant Conic (X<sub>0</sub>=-96.0, Y<sub>0</sub>=40.0, Y<sub>1</sub>=20.0, Y<sub>2</sub>=60.0)

*Datum:* NAD83

*Spheroid:* GRS80

*Derived Surrogates:* 801, 820

### **2.3.3 Gulf of Mexico Platforms**

ERG, Inc. estimated the locations of class 1 and 2 commercial shipping vessels in the Gulf of Mexico and associated these locations with the density of shipping activity surrounding oil and gas platforms (Figure 3). A Shapefile that defines the fraction of total Gulf activity density around the energy development platforms was used to create a Gulf Shipping Lanes surrogate.

The attributes in the file used to compute surrogates include:

- Fraction – fraction of Gulf-wide activity associated with individual oil and gas platforms

*Shapefile Name:* GulfofMexico\_SupportVessels\_Density\_WRF

*Shapefile Type:* Polygon

*Year:* 2010

*Spatial Coverage:* Gulf of Mexico

*Attributes:* GRIDCODE, Area, Fraction

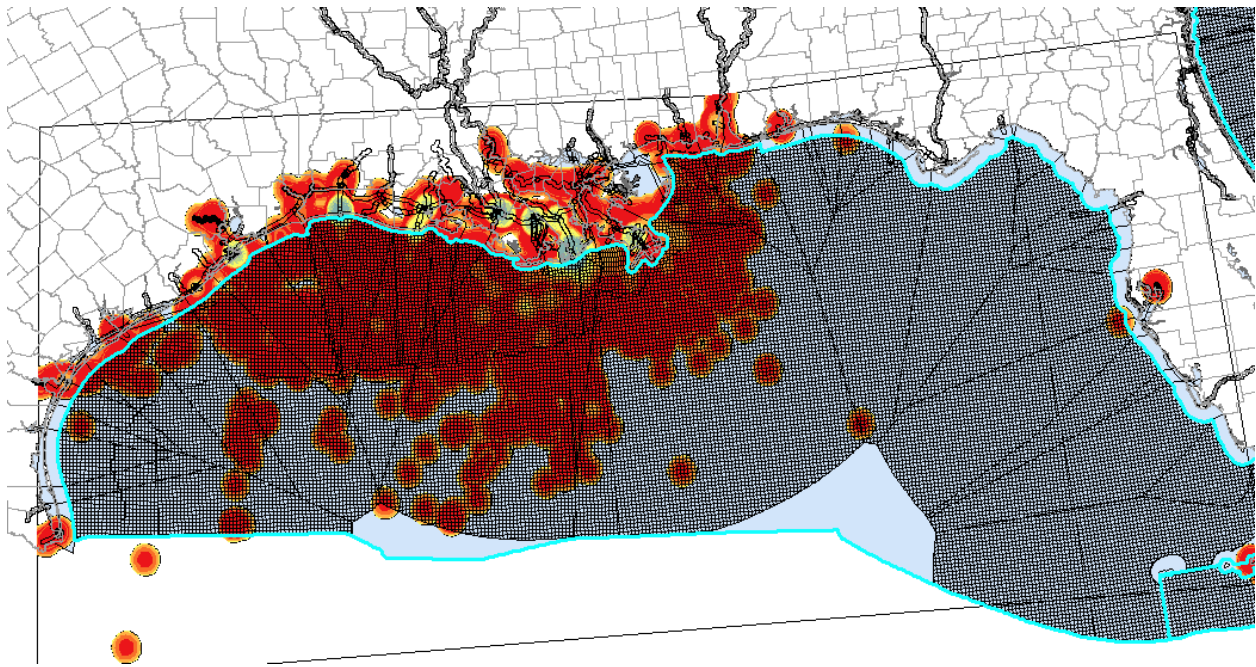
*Projection:* Lambert Conformal Conic (X<sub>0</sub>=-97.0, Y<sub>1</sub>=33.0, Y<sub>2</sub>=45.0, Y<sub>0</sub>=40.0)

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters



*Derived Surrogates: 802*



**Figure 3. Gulf of Mexico oil and gas platform support vessel densities**

### 2.3.4 Offshore Commercial Shipping

Offshore shipping lane areas for the Pacific, Atlantic, Gulf of Mexico, and Great Lakes weighted by activity from the NEI2014.

The attributes in the file used to compute surrogates include:

- Area\_sqmi – port areas in miles<sup>2</sup>
- ACTIVITYKW – annual KW activities for C1/C2 vessels from the NEI2014

*Shapefile Name:* ShippingLanes\_2014NEI

*Shapefile Type:* Polygon

*Year:* 2014

*Spatial Coverage:* Conterminous U.S.

*Attributes:* Area\_sqmi, ShapeID, FIPS, ActivityKW

*Projection:* Equidistant Conic (X<sub>0</sub>=-96.0, Y<sub>0</sub>=40.0, Y<sub>1</sub>=20.0, Y<sub>2</sub>=60.0)

*Datum:* NAD83

*Spheroid:* GRS80

*Derived Surrogates:* 802, 805, 806

## 2.4 FEMA Building Footprints

HAZUS is a nationally standardized methodology that contains models for estimating potential losses from natural disasters. A key component of the HAZUS database is information on building locations, sizes, and classifications. These geospatial data on buildings are useful for allocating inventory data to modeling grids because they contain details on the locations and sizes of different building types. The Federal Emergency Management Agency HAZUS-MH version 2 was released in September 2011 and contains square footage data for different types of buildings throughout the U.S. The building square footage data are used to identify building classifications (i.e. commercial, residential, industrial, institutional) for allocating non-point inventory sources to modeling grids. UNC-IE merged Census-block level data in state Shapefiles into a national Shapefile for select database attributes.

The attributes in the file used to compute surrogates include:

- COM# - commercial building square footage
- IND# - industrial building square footage
- RES# - residential building square footage
- EDU# - educational building square footage
- REL# - religious building square footage
- GOV# - government building square footage

*Shapefile Name:* fema\_bsf\_2002bnd

*Shapefile Type:* Polygon

*Year:* 2011

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* COM#, IND#, RES#, EDU#, REL#, GOV#

*Projection:* Lambert Conformal Conic ( $X_0=-97.0$ ,  $Y_1=33.0$ ,  $Y_2=45.0$ ,  $Y_0=40.0$ )

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 500, 505, 506, 510, 512, 515, 520, 525, 526, 527, 535, 540, 545, 555, 560, 575, 580, 585, 590, 595, 596

## 2.5 Landcover Surrogates

2011 National Land Cover Database (NLCD) land cover data were used to create this shapefile. UNC-IE first dissolved the original 30-meter resolution NLCD raster data to 500-meter resolution using the dominant land use class. The raster data were then converted to a vector polygon file and a spatial join was performed to combine the Continental U.S. and Alaska land cover data into a single shapefile. The attribute "GRIDCODE" in the Shapefile is used to identify

the different landuse categories. The surrogates for different land cover categories were generated by filtering the data based on the GRIDCODE.

The attributes in the file used to compute surrogates are listed in Table 5.

**Table 5. NLCD 2011 surrogate attributes**

GRIDCODE	NLCD categories	Surrogate ID	Surrogate Name
22	Developed, Low Intensity	300	NLCD Low Intensity Development
23	Developed, Medium Intensity	301	NLCD Med Intensity Development
24	Developed, High Intensity	302	NCLD High Intensity Development
21	Developed, Open Space	303	NLCD Open Space
21, 22		304	NLCD Open + Low
22, 23		305	NLCD Low + Med
23, 24		306	NLCD Med + High
21, 22, 23, 24		307	NLCD All Development
22, 23, 24		308	NLCD Low + Med + High
21, 22, 23		309	NLCD Open + Low + Med
81, 82		310	NLCD Total Agriculture
81	Pasture/Hay	318	NLCD Pasture Land
82	Cultivated Crops	319	NLCD Crop Land
41, 42, 43	Deciduous Forest (41), Evergreen Forest (42), Mixed Forest (43)	320	NLCD Forest Land
21, 31, 41, 42, 43, 52, 71	Developed, Open Space (21), Barren Land (31), Forest, Shrub/Scrub, Grassland	321	NLCD Recreational Land
! = 11	Everything but water	340	Land
11	Water	350	Water

*Shapefile Name:* CONUS\_AK\_NLCD\_2011\_500m\_WGS

*Shapefile Type:* Polygon

*Year:* 2011

*Spatial Coverage:* Conterminous U.S.

*Attributes:* GRIDCODE

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* See Table 5

## **2.6 Oil and Gas Production**

### **2.6.1 Eastern Research Group, Inc. Oil and Gas Production Data (HPDI Database)**

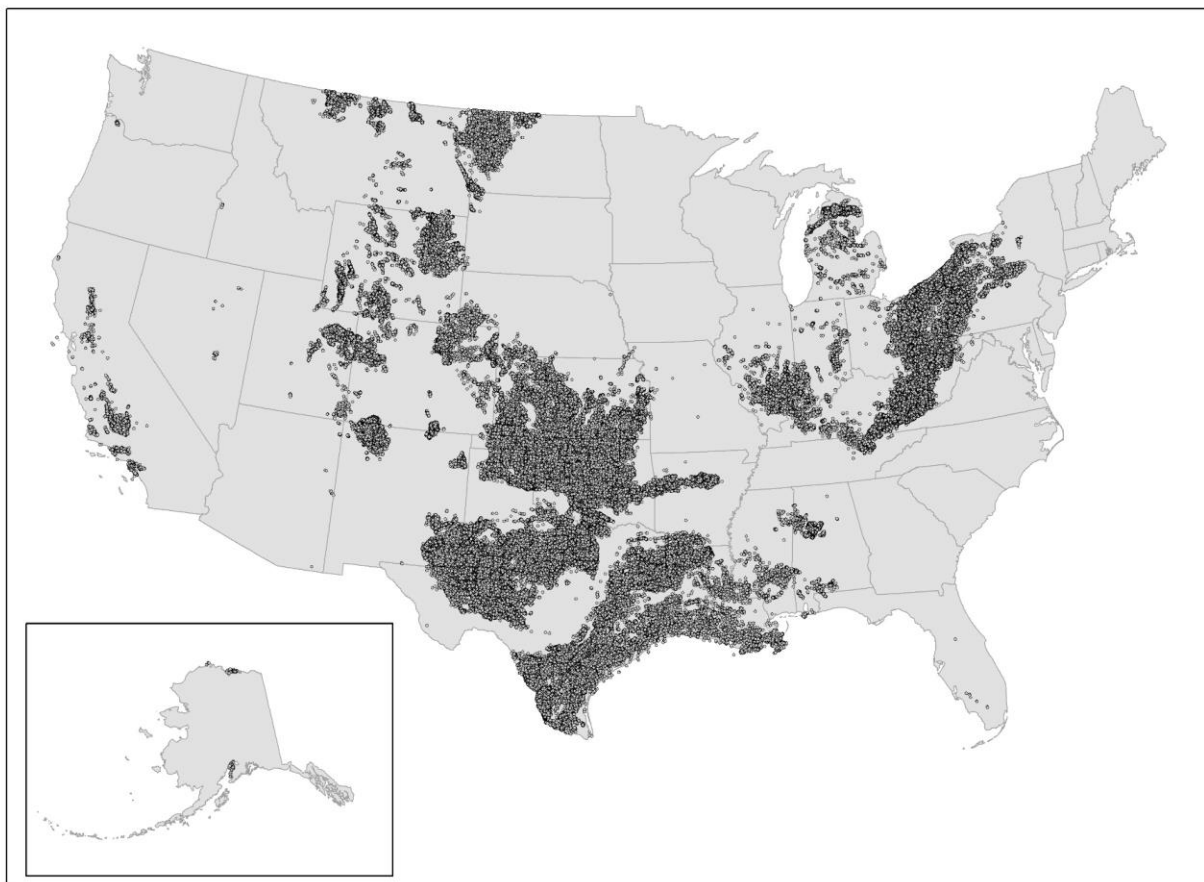
The primary activity data source that ERG used for developing oil and gas spatial surrogates were data from Drilling Info (DI) Desktop's HPDI database (ERG, 2016). This database contains well-level location, production, and exploration statistics at the monthly level. Due to a proprietary agreement with DI Desktop, individual well locations and ancillary production data cannot be made publicly available, but aggregated statistics are allowed. The HPDI data represented approximately 94% of the activity data used for these surrogates.

For the remaining 6% of the activity, ERG supplemented the HPDI with additional data from Oil and Gas Commission (OGC) websites. In many cases, the correct surrogate parameter was not available (e.g., feet drilled), but an alternative surrogate parameter was available (e.g., number of spudded wells) and used for generating the spatial surrogates.

In total, ERG compiled over 1.43 million unique well locations from the above data sources. The well locations cover 34 states and 1,158 counties. Each well was uploaded into ArcGIS, and assigned to 2-km and 4-km Continental U.S. grid identifiers. Well locations are presented in Figure 4.

As the HPDI data collected by ERG are proprietary and cannot be redistributed, a set of de-identified Shapefiles that include the number of wells or production units in 2-km to 4-km grid cells (polygons) will be made available to the public. The surrogates developed from these public Shapefiles will use the number of production units in each grid cell as the weighting attribute. Although it will not be possible to exactly reproduce the oil and gas production surrogates in the EPA modeling platform without the proprietary Shapefile data, the de-identified production Shapefiles will provide a close approximation of these data to use for the development of surrogates on customized modeling domains.

For additional details on these Shapefiles and spatial surrogates see ERG (2016).



**Figure 4. 2015 active oil and gas well locations.**

The shapefile names and attributes in the ERG/HPDI files used to compute the surrogates are listed in Table 6.

**Table 6. HPMS Oil and gas surrogate codes**

Surrogate Name	Surrogate ID	Shapefile Name	Weight Attribute
Spud Count - CBM Wells	670	SpudCount_CBM	SPUD_COU_1
Spud Count - Gas Wells	671	SpudCount_Gas	SPUD_COU_1
Gas Production at Oil Wells	672	AssocGasProduction	ASSOCIAT_1
Oil Production at CBM Wells	673	CondensateCBMProd	CONDENSA_1
Unconv. Well Completion Counts	674	SpudCount_HF	SPUD_HF_21
Well Count - All Producing	676	AllProductionWells	TOTAL_PR_1
Well Count - All Exploratory	677	AllExploratoryWells	TOTAL_EX_1
Completions at Gas Wells	678	Completions_Gas	COMPLETI_1
Completions at CBM Wells	679	Completions_CBM	COMPLETI_1
Spud Count - Oil Wells	681	SpudCount_Oil	SPUD_COU_1
Produced Water at All Wells	683	ProducedWater_All	PRODUCED_1
Completions at Oil Wells	685	Completions_Oil	COMPLETI_1
Completions at All Wells	686	Completions_All	COMPLETI_1
Feet Drilled at All Wells	687	FeetDrilled_All	FEET_DRI_1

Well Counts - CBM Wells	691	CBMWell_Counts	CBM_WELLS1
Spud Count - All Wells	692	SpudCount_All	SPUD_COU_1
Well Count - All Wells	693	AllWells	TOTAL_WE_1
Oil Production at Oil Wells	694	OilProduction	OIL_PROD_1
Well Count - Oil Wells	695	OilWell_Counts	OIL_WELL_1
Gas Production at Gas Wells	696	GasProduction	GAS_PROD_1
Oil Production at Gas Wells	697	CondensateGasProd	CONDENSA_1
Well Count - Gas Wells	698	GasWell_Counts	CBM_WELLS1
Gas Production at CBM Wells	699	CBMProduction	CBM_PROD_1

*Type:* Polygon

*Year:* 2015

*Spatial Coverage:* Conterminous U.S.

*Attributes:* The attributes are different for each Shapefile. See ERG (2016) for details about the Shapefiles and the derived surrogates

*Projection:* Lambert Conformal Conic ( $X_0=-97.0$ ,  $Y_1=33.0$ ,  $Y_2=45.0$ ,  $Y_0=40.0$ )

*Datum:* Unknown

*Spheroid:* Unknown

*Derived Surrogates:* 670-699

## 2.7 Other Industrial and Commercial Activities

### 2.7.1 Refineries and Tank Farms

Energy Information Administration 2015/16 locations of downstream liquid petroleum processing and storage facilities. UNC-IE downloaded and joined (spatial) the following datasets from the EIA website<sup>6</sup>: Product Terminals, Refineries, LNG Import/Export Terminals, and Crude Oil Rail Terminals.

The attributes in the file used to compute surrogates include:

- NONE – point locations of the commodities used to derive the surrogate

*Shapefile Name:* EIA\_2015\_US-Oil

*Shapefile Type:* Point

*Year:* 2015-16

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* Site\_Name,CITY,STATE,PADD,STATION\_TY,FACILITY\_T,HANDLING, LONGITUDE,LATITUDE,Source,Company,Corp,Site,AD\_Mbpd,VDist\_Mbpd,CaDis\_Mbpd,V

---

<sup>6</sup> [https://www.eia.gov/maps/layer\\_info-m.php](https://www.eia.gov/maps/layer_info-m.php)

Redu\_Mbpd,CaRef\_Mbpd,Isal\_Mbpd,HDS\_Mbpd,Cokin\_Mbpd,Asph\_Mbpd,Period\_,Data\_peri  
o,Facility,Owner,Operator,County,ImpExp,Period,Cap\_Bcfd,Liq\_Bcfd,Stora\_MMcf,

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* 650

### **2.7.2 Mines**

2011 USGS Mineral Resources Data System (MRDS) point locations of metallic and nonmetallic mineral resources.

The attributes in the file used to compute surrogates include:

- DEV\_STAT: development status; DEV\_STAT = producer was used for the surrogate calculation

*Shapefile Name:* USGS\_2011\_mines

*Shapefile Type:* Point

*Year:* 2011

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* DEP\_ID, SITE\_NAME,DEV\_STAT,URL,CODE\_LIST,

*Projection:* Geographic

*Datum:* WGS84

*Spheroid:* WGS84

*Derived Surrogates:* 860

### **2.7.3 Commercial Timber**

2016 U.S. Forest Service activities accomplished as a part of the timber harvest program of work. It is important to note that this layer does not contain all timber harvest activities because the spatial portion of the activity description is not required and is self-reported by Forest Service units

The attributes in the file used to compute surrogates include:

- NONE – locations of the harvest activities were used to generate the surrogate

*Shapefile Name:* USFS\_2016\_TimberHarvest

*Shapefile Type:* Polygon

*Year:* 2016

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* ADMIN\_FORE,ADMIN\_REGL,ADMIN\_FO\_1,PROCLAIMED,ADMIN\_DIST,  
ADMIN\_DI\_1,HOME\_ORG,ACTIVITY\_U,SUID,FACTS\_ID,SUBUNIT,SUBUNIT\_CN,SUB

UNIT\_NA,SUBUNIT\_SI,SALE\_NAME,ACTIVITY\_C,ACTIVITY\_1,ACTIVITY\_2,ACTIVITY\_N,TREATMENT\_,NBR\_UNITS\_,UOM,NBR\_UNITS1,DATE\_PLANN,DATE\_ACCOM,DATE\_COMPL,FY\_PLANNED,FY\_ACCOMPL,FY\_COMPLET,FUND\_CODES,COST\_PER\_U,NEPA\_PROJE,NEPA\_DOC\_N,NEPA\_PRO\_1,METHOD\_COD,METHOD\_DES,EQUIPMENT\_,EQUIPMENT1,IMPLEMENTA,IMPLEMEN\_1,IMPLEMEN\_2,WORK\_AGENT,LAND\_SUITA,LAND\_SUI\_1,PRODUCTIVI,PRODUCTI\_1,OWNERSHIP\_,OWNERSHIP1,ASPECT\_,ELEVATION,SLOPE,STATE\_ABBR,WATERSHED\_,SUBUNIT\_UO,STAGE,STAGE\_DESC,DATA\_SOURC,DATA\_SOU\_1,ACCURACY,CRC\_VALUE,UK,EDW\_INSERT,ETL\_MODIFIED,REV\_DATE,GIS\_ACRES,SHAPE\_AREA,SHAPE\_LEN

*Projection:* Geographic

*Datum:* NAD83

*Spheroid:* GRS80

*Derived Surrogates:* 860

#### 2.7.4 Golf Courses

The Point of Interest (POI) Factory is a crowd-sourced GIS data resource. The GolfCourses-USA Shapefile was submitted on June 7, 2010 and is updated continually by the POI user community. The dataset was compiled from the website locator on [www.golflinks.com](http://www.golflinks.com). Some of the coordinates were obtained using the Google Earth streetview and GPS visualizer.

The attributes in the file used to compute surrogates include:

- NONE – the surrogate was generated using the point locations of the golf courses

*Shapefile Name:* POI-Factory\_2015\_GolfCourses

*Shapefile Type:* Point

*Year:* 2015

*Spatial Coverage:* All U.S. states, no territories

*Attributes:* lon,lat,name,type,address,city,state,phone

*Projection:* Lambert Conformal Conic ( $X_0=-97.0$ ,  $Y_1=33.0$ ,  $Y_2=45.0$ ,  $Y_0=40.0$ )

*Datum:* NAD83

*Spheroid:* GRS80

*Derived Surrogates:* 850

#### 2.7.5 Airport Point Emissions

Locations of U.S. airports weighted by NEI 2008 total emissions

The attributes in the file used to compute surrogates include:

- TOTAL\_EMIS – total emissions of all sources at each airport location reported in the NEI2008v2 inventory

*Shapefile Name:* US\_Airports\_NEI08v2\_WRF



*Shapefile Type:* Point

*Year:* 2008

*Spatial Coverage:* Conterminous U.S.

*Attributes:* FIPS,NAME,LAT,LON,TOTAL\_EMIS,YEAR

*Projection:* Lambert Conformal Conic ( $X_0=-97.0$ ,  $Y_1=33.0$ ,  $Y_2=45.0$ ,  $Y_0=40.0$ )

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,000 meters

*Derived Surrogates:* 710

## 2.7.6 Airport Area

Spatial areas of U.S. airports in 199.

The attributes in the file used to compute surrogates include:

- TOTAL\_EMIS – total emissions of all sources at each airport location reported in the NEI2008v2 inventory

*Shapefile Name:* airport\_area

*Shapefile Type:* Polygon

*Year:* 1999

*Spatial Coverage:* Conterminous U.S.

*Attributes:* AREA,PERIMETER,AIRPORT\_AR,AIRPORT\_\_1,NAME,FCC,LOC\_ID,USE,OWNERNAME,FIPSSTCO,STATE,COUNTY

*Projection:* Lambert Conformal Conic ( $X_0=-97.0$ ,  $Y_1=33.0$ ,  $Y_2=45.0$ ,  $Y_0=40.0$ )

*Datum:* NAD83

*Spheroid:* Normal Sphere with radius 6,370,997 meters

*Derived Surrogates:* 711

## 3 Known Issues and Future Work

While the Shapefiles and surrogates documented here and in the accompanying spreadsheet represent the best available information for allocating non-point emissions to modeling grids, there are some caveats to these data and issues that should be addressed in the future.

- The total agriculture surrogate (310) is not specific enough for livestock sources and could be improved using a point Shapefile of confined animal feeding operations (CAFOs). Although we were not able to find a national Shapefile of CAFO locations, these data are available from some states. State departments of water quality track CAFO locations for monitoring runoff and water permitting purposes. It will be worthwhile to collect data from as many states as possible for developing a national CAFO surrogate to use for allocating non-point livestock inventories to model grids.

- The mines surrogate (860) is missing many of the coal mines in the U.S. Shapefiles available from state mining commissions are a better source of coal mining data and should be used to augment the national data.
- Data are available from LADCO on shipping lanes in the Great Lakes. These data should be used in the calculation of the commercial shipping (802) surrogate.

## 4 Surrogate Revision History

## 5 REFERENCES

- Bergin, M. et al (2009) ERTAC Rail: A collaborative effort in building a railroad-related emission inventory between Eastern states' air protection agencies and participation with the railroad industry, presented at the 18<sup>th</sup> Annual Emission Inventory Conference, U.S. EPA, Baltimore, MD, <http://www.epa.gov/ttnchie1/conference/ei18/session6/bergin.pdf>.
- ERG (2016) Technical Memorandum: Modeling Allocation Factors for 2014 Oil and Gas Nonpoint Tool, To: Alison Eyth and Laurie Trinca, US EPA OAQPS, From: Regi Oomen et al., ERG, September 2, 2016.
- Eyth, A. et al. (2007) Accounting for Land Use Changes in Projecting Future-Year Emissions Scenarios, presented at the 16<sup>th</sup> Annual Emission Inventory Conference, U.S. EPA, Raleigh, NC, <http://www.epa.gov/ttnchie1/conference/ei16/session7/eyth.pdf>.
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K. (2015) [Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information](#). *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354.
- U.S. Bureau of Transportation Statistics (BTS) (2014) 2014 National Transportation Atlas Database Liner Notes. [http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national\\_transportation\\_atlas\\_database/2014/liner](http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national_transportation_atlas_database/2014/liner).
- U.S. Census Bureau (2014) 2014 TIGER/Line Shapefiles Technical Documentation. [http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2014/TGRSHP2014\\_TechDoc.pdf](http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2014/TGRSHP2014_TechDoc.pdf)
- U.S. Department of Transportation (DOT) (2013) Highway Performance Monitoring System Field Manual. <https://www.fhwa.dot.gov/policyinformation/hpms/fieldmanual/>
- U.S. EPA (2016) Population and Activity of On Road Vehicles in MOVES2014. EPA-420-R-16-003. January 2016.
- USGS (2005) Mineral Resources Data System, edition 20160305. <http://mrdata.usgs.gov/mrds/>, Reston, VA.
- UNC-IE (2014) Emissions Modeling Platform Spatial Surrogate Documentation, Prepared for Rich Mason, US EPA OAQPS Contract No. EP-D-12-044 Work Assignment 2-02, UNC-EMAQ(2-02)-011.v2, Chapel Hill, NC, September 30, 2014.



## Appendix A – Surrogate Specifications

**Table 7. Primary surrogate specifications**

ID	Surrogate Name	Weight Shapefile	Attribute	Weight/Filter Function
100	Population	ACS_2014_5YR_PopHousing	POP2014	
110	Housing	ACS_2014_5YR_PopHousing	HU2014	
131	Urban Housing	ACS_2014_5YR_PopHousing	NONE	HUD_CODE=1
132	Suburban Housing	ACS_2014_5YR_PopHousing	NONE	HUD_CODE=1
134	Rural Housing	ACS_2014_5YR_PopHousing	NONE	HUD_CODE=4
137	Housing Change	ACS_2014_5YR_PopHousing	HUCH14_10	
140	Housing Change and Population	NA	NA	0.5*Housing Change+0.5*Population
150	Residential Heating - Natural Gas	ACS_2014_5YR_PopHousing	UTIL_GAS	
160	Residential Heating - Wood	ACS_2014_5YR_PopHousing	WOOD	
170	Residential Heating - Distillate Oil	ACS_2014_5YR_PopHousing	FUEL_OIL	
180	Residential Heating - Coal	ACS_2014_5YR_PopHousing	COAL	
190	Residential Heating - LP Gas	ACS_2014_5YR_PopHousing	LP_GAS	
201	Urban Restricted Road Miles		NONE	RDTYPE=04
202	Urban Restricted AADT		aadt	RDTYPE=04
205	Extended Idle Locations	2016_PIL_16Aug	trucks_med	
211	Rural Restricted Road Miles		NONE	RDTYPE=02
212	Rural Restricted AADT		aadt	RDTYPE=02
221	Urban Unrestricted Road Miles		NONE	RDTYPE=05
222	Urban Unrestricted AADT		aadt	RDTYPE=05
231	Rural Unrestricted Road Miles		NONE	RDTYPE=03
232	Rural Unrestricted AADT		aadt	RDTYPE=03
239	Total AADT		aadt	RDTYPE=02,03,04,05
240	Total Road Miles		NONE	RDTYPE=02,03,04,05

241	Total Restricted Road Miles		NONE	RDTYPE=02,04
242	All Restricted Road AADT		aadt	RDTYPE=02,04
243	Total Unrestricted Road Miles		NONE	RDTYPE=03,05
244	All Unrestricted Road AADT		aadt	RDTYPE=03,05
258	Intercity Bus Terminals	NTAD_2014_ipcd	NONE	BUS_INTERC = 1
259	Transit Bus Terminals	NTAD_2014_ipcd	NONE	BUS_TRANSI = 1
260	Total Railroad Miles	TIGER_2014_Rail	NONE	
261	NTAD Total Railroad Density	NTAD_2014_Rail	DENS	RAILTYPE=1,2,3
271	NTAD Class 1 2 3 Railroad Density	NTAD_2014_Rail	DENS	RAILTYPE="1"
272	NTAD Amtrak Railroad Density	TIGER_2014_Rail	DENS	RAILTYPE="2"
273	NTAD Commuter Railroad Density	NTAD_2014_Rail	DENS	RAILTYPE="3"
275	ERTAC Rail Yards	ERTAC_railyard_WRF	NOXPMTOT	
280	Class 2 and 3 Railroad Miles	TIGER_2014_Rail	NONE	MTFCC=R1051,R1052
300	NLCD Low Intensity Development	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=22
301	NLCD Med Intensity Development	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=23
302	NLCD High Intensity Development	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=24
303	NLCD Open Space	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=21
304	NLCD Open + Low	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=21,22
305	NLCD Low + Med	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=22,23
306	NLCD Med + High	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=23,24
307	NLCD All Development	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=21,22,23,24
308	NLCD Low + Med + High	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=22,23,24
309	NLCD Open + Low + Med	CONUS_AK_NLCD_2011_500m_WGS		GRIDCODE=21,22,23
310	NLCD Total Agriculture	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=81,82
318	NLCD Pasture Land	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=81
319	NLCD Crop Land	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=82
320	NLCD Forest Land	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=41,42,43

321	NLCD Recreational Land	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=21,31,41,42,43,52,71
340	NLCD Land	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE!=11
350	NLCD Water	CONUS_AK_NLCD_2011_500m_WGS	NONE	GRIDCODE=11
500	Commercial Land	fema_bsf_2002bnd	NA	
505	Industrial Land	fema_bsf_2002bnd	NA	
506	Education	fema_bsf_2002bnd	NA	
507	Heavy Light Construction Industrial Land	fema_bsf_2002bnd	NA	
510	Commercial plus Industrial	fema_bsf_2002bnd	NA	
515	Commercial plus Institutional Land	fema_bsf_2002bnd	NA	
520	Commercial plus Industrial plus Institutional	fema_bsf_2002bnd	NA	
525	Golf Courses plus Institutional plus Industrial plus Commercial	NA	NA	0.5*Commercial plus Industrial plus Institutional+0.5*Golf Courses
526	Residential - Non-Institutional	fema_bsf_2002bnd	NA	
527	Single Family Residential	fema_bsf_2002bnd	RES1	
535	Residential + Commercial + Industrial + Institutional + Government	fema_bsf_2002bnd	NA	
540	Retail Trade (COM1)	fema_bsf_2002bnd	COM1	
545	Personal Repair (COM3)	fema_bsf_2002bnd	COM3	
555	Professional/Technical (COM4) plus General Government (GOV1)	fema_bsf_2002bnd	NA	
560	Hospital (COM6)	fema_bsf_2002bnd	COM6	
575	Light and High Tech Industrial (IND2 + IND5)	fema_bsf_2002bnd	NA	
580	Food, Drug, Chemical Industrial (IND3)	fema_bsf_2002bnd	IND3	
585	Metals and Minerals Industrial (IND4)	fema_bsf_2002bnd	IND4	
590	Heavy Industrial (IND1)	fema_bsf_2002bnd	IND1	

595	Light Industrial (IND2)	fema_bsf_2002bnd	IND2	
596	Industrial plus Institutional plus Hospitals	fema_bsf_2002bnd		
650	Refineries and Tank Farms	EIA_2015_US-Oil	NONE	
670	Spud Count - CBM Wells	SpudCount_CBM	SPUD_COU_1	
671	Spud Count - Gas Wells	SpudCount_Gas	SPUD_COU_1	
672	Gas Production at Oil Wells	AssocGasProduction	ASSOCIAT_1	
673	Oil Production at CBM Wells	CondensateCBMProd	CONDENSA_1	
674	Unconventional Well Completion Counts	SpudCount_HF	SPUD_HF_21	
676	Well Count - All Producing	AllProductionWells	TOTAL_PR_1	
677	Well Count - All Exploratory	AllExploratoryWells	TOTAL_EX_1	
678	Completions at Gas Wells	Completions_Gas	COMPLETI_1	
679	Completions at CBM Wells	Completions_CBM	COMPLETI_1	
681	Spud Count - Oil Wells	SpudCount_Oil	SPUD_COU_1	
683	Produced Water at All Wells	ProducedWater_All	PRODUCED_1	
685	Completions at Oil Wells	Completions_Oil	COMPLETI_1	
686	Completions at All Wells	Completions_All	COMPLETI_1	
687	Feet Drilled at All Wells	FeetDrilled_All	FEET_DRI_1	
691	Well Counts - CBM Wells	CBMWell_Counts	CBM_WELLS1	
692	Spud Count - All Wells	SpudCount_All	SPUD_COU_1	
693	Well Count - All Wells	AllWells	TOTAL_WE_1	
694	Oil Production at Oil Wells	OilProduction	OIL_PROD_1	
695	Well Count - Oil Wells	OilWell_Counts	OIL_WELL_1	
696	Gas Production at Gas Wells	GasProduction	GAS_PROD_1	
697	Oil Production at Gas Wells	CondensateGasProd	CONDENSA_1	
698	Well Count - Gas Wells	GasWell_Counts	CBM_WELLS1	

699	Gas Production at CBM Wells	CBMProduction	CBM_PROD_1	
710	Airport Points	US_Airports_NEI08v2_WRF	TOTAL_EMIS	
711	Airport Areas	airport-area	AREA	
801	Port Areas	Ports_2014NEI	Area_sqmi	
802	Commercial Shipping			
803	Gulf Shipping Lanes			
805	Offshore Shipping Area	ShippingLanes_2014NEI	Area_sqmi	
806	Offshore Shipping NEI2014 Activity	ShippingLanes_2014NEI	ACTIVITYKW	
807	Navigable Waterway Miles	NTAD_2014_Waterway	LENGTH	
820	Ports NEI2014 Activity	Ports_2014NEI	ACTIVITYKW	
850	Golf Courses	POI-Factory_2015_GolfCourses	NONE	
860	Mines	USGS_2011_mines	NONE	DEV_STAT=Producer
890	Commercial Timber	USFS_2016_TimberHarvest	NONE	

**Table 8. Surrogate gapfilling specifications**

ID	Surrogate Name	Secondary	Tertiary	Quaternary
100	Population			
110	Housing	Population		
131	Urban Housing	Housing	Population	NLCD Land
132	Suburban Housing	Housing	Population	NLCD Land
134	Rural Housing	Housing	Population	NLCD Land
137	Housing Change	Housing	Population	NLCD Land
140	Housing Change and Population	Population		
150	Residential Heating - Natural Gas	Housing		
160	Residential Heating - Wood	Housing		
170	Residential Heating - Distillate Oil	Housing		
180	Residential Heating - Coal	Housing		
190	Residential Heating - LP Gas	Housing		
201	Urban Restricted Road Miles	Total Road Miles		
202	Urban Restricted AADT	Total Road Miles		
205	Extended Idle Locations	Total Road Miles		



211	Rural Restricted Road Miles	Total Road Miles		
212	Rural Restricted AADT	Total Road Miles		
221	Urban Unrestricted Road Miles	Total Road Miles		
222	Urban Unrestricted AADT	Total Road Miles		
231	Rural Unrestricted Road Miles	Total Road Miles		
232	Rural Unrestricted AADT	Total Road Miles		
239	Total AADT	Total Road Miles		
240	Total Road Miles			
241	Total Restricted Road Miles	Total Road Miles		
242	All Restricted Road AADT	Total Road Miles		
243	Total Unrestricted Road Miles	Total Road Miles		
244	All Unrestricted Road AADT	Total Road Miles		
258	Intercity Bus Terminals	Commercial plus Industrial	Population	NLCD Land
259	Transit Bus Terminals	Commercial plus Industrial	Population	NLCD Land
260	Total Railroad Miles	Total Road Miles	Population	
261	NTAD Total Railroad Density	Total Railroad Miles	Total Road Miles	Population
271	NTAD Class 1 2 3 Railroad Density	NTAD Total Railroad Density	Total Railroad Miles	Total Road Miles
272	NTAD Amtrak Railroad Density	NTAD Total Railroad Density	Total Railroad Miles	Total Road Miles
273	NTAD Commuter Railroad Density	NTAD Total Railroad Density	Total Railroad Miles	Total Road Miles
275	ERTAC Rail Yards	NTAD Total Railroad Density	Total Railroad Miles	Population
280	Class 2 and 3 Railroad Miles	Total Railroad Miles	Total Road Miles	Population
300	NLCD Low Intensity Development	Rural Housing	Population	NLCD Land
301	NLCD Med Intensity Development	Suburban Housing	Population	NLCD Land
302	NLCD High Intensity Development	Population	NLCD Land	
303	NLCD Open Space	Rural Housing	Population	NLCD Land
304	NLCD Open + Low	Suburban Housing	Population	NLCD Land
305	NLCD Low + Med	Urban Housing	Population	NLCD Land
306	NLCD Med + High	Urban Housing	Population	NLCD Land
307	NLCD All Development	Urban Housing	Population	NLCD Land
308	NLCD Low + Med + High	Urban Housing	Population	NLCD Land
309	NLCD Open + Low + Med	Rural Housing	NLCD Land	

310	NLCD Total Agriculture	Rural Housing	NLCD Land	
318	NLCD Pasture Land	Rural Housing	NLCD Land	
319	NLCD Crop Land	Rural Land Area	NLCD Land	
320	NLCD Forest Land	Rural Land Area	NLCD Land	
321	NLCD Recreational Land			
340	NLCD Land			
350	NLCD Water	Population	NLCD Land	
500	Commercial Land	Population	NLCD Land	
505	Industrial Land	Population	NLCD Land	
506	Education	Population	NLCD Land	
507	Heavy Light Construction Industrial Land	Population	NLCD Land	
510	Commercial plus Industrial	Population	NLCD Land	
515	Commercial plus Institutional Land	Population	NLCD Land	
520	Commercial plus Industrial plus Institutional	Population	NLCD Land	
525	Golf Courses plus Institutional plus Industrial plus Commercial	Golf Courses	Commercial plus Industrial plus Institutional	Population
526	Residential - Non-Institutional	Housing	Population	
527	Single Family Residential	Housing	Population	NLCD Land
535	Residential + Commercial + Industrial + Institutional + Government	Population	NLCD Land	
540	Retail Trade (COM1)	Personal Repair (COM3)	Commercial Land	Population
545	Personal Repair (COM3)	Commercial Land	Population	NLCD Land
555	Professional/Technical (COM4) plus General Government (GOV1)	Commercial Land	Residential + Commercial + Industrial + Institutional + Government	Population
560	Hospital (COM6)	Commercial Land	Population	
575	Light and High Tech Industrial (IND2 + IND5)	Industrial Land	Population	NLCD Land
580	Food, Drug, Chemical Industrial (IND3)	Industrial Land	Population	NLCD Land
585	Metals and Minerals Industrial (IND4)	Industrial Land	Population	NLCD Land
590	Heavy Industrial (IND1)	Industrial Land	Population	NLCD Land
595	Light Industrial (IND2)	Industrial Land	Population	NLCD Land
596	Industrial plus Institutional plus Hospitals	Population	NLCD Land	
650	Refineries and Tank Farms	Industrial Land	Population	NLCD Land

670	Spud Count - CBM Wells	Industrial Land	Population	NLCD Land
671	Spud Count - Gas Wells	Industrial Land	Population	NLCD Land
672	Gas Production at Oil Wells	Industrial Land	Population	NLCD Land
673	Oil Production at CBM Wells	Industrial Land	Population	NLCD Land
674	Unconventional Well Completion Counts	Industrial Land	Population	NLCD Land
676	Well Count - All Producing	Industrial Land	Population	NLCD Land
677	Well Count - All Exploratory	Industrial Land	Population	NLCD Land
678	Completions at Gas Wells	Industrial Land	Population	NLCD Land
679	Completions at CBM Wells	Industrial Land	Population	NLCD Land
681	Spud Count - Oil Wells	Industrial Land	Population	NLCD Land
683	Produced Water at All Wells	Industrial Land	Population	NLCD Land
685	Completions at Oil Wells	Industrial Land	Population	NLCD Land
686	Completions at All Wells	Industrial Land	Population	NLCD Land
687	Feet Drilled at All Wells	Industrial Land	Population	NLCD Land
691	Well Counts - CBM Wells	Industrial Land	Population	NLCD Land
692	Spud Count - All Wells	Industrial Land	Population	NLCD Land
693	Well Count - All Wells	Industrial Land	Population	NLCD Land
694	Oil Production at Oil Wells	Industrial Land	Population	NLCD Land
695	Well Count - Oil Wells	Industrial Land	Population	NLCD Land
696	Gas Production at Gas Wells	Industrial Land	Population	NLCD Land
697	Oil Production at Gas Wells	Industrial Land	Population	NLCD Land
698	Well Count - Gas Wells	Industrial Land	Population	NLCD Land
699	Gas Production at CBM Wells	Industrial Land	Population	NLCD Land
710	Airport Points	Population	NLCD Land	
711	Airport Areas	Airport Points	Population	NLCD Land
801	Port Areas	Navigable Waterway Miles	NLCD Water	
802	Commercial Shipping			
803	Gulf Shipping Lanes			
805	Offshore Shipping Area			
806	Offshore Shipping NEI2014 Activity	Offshore Shipping Area	Navigable Waterway Miles	NLCD Water
807	Navigable Waterway Miles	NLCD Water		
820	Ports NEI2014 Activity	Port Areas	Navigable Waterway Miles	NLCD Water

850	Golf Courses	Housing	Population	NLCD Land
860	Mines	Rural Housing	NLCD Land	
890	Commercial Timber	NLCD Forest Land	Rural Housing	Population

## Appendix B – 12km Surrogate Plots

